Jordan Love

I. Motivation

The Boston waterfront is one of the most beautiful, and sought-after areas of the city. This past year, however, Boston's number one up and coming area was temporarily reduced to little more than a frozen pond due to high flood waters brought in by brutal storms. This flooding was reported as the worst Boston has seen in years, however the city of Boston predicts that it will only get worse in the years to come. Fortunately, the waterfront is not a cheap place to live, and those who dwell there aren't likely to have much difficulty relocating if the rising waters become a bigger issues. The waterfront wasn't the only area affected, however. While the effect to areas more inland was small, it was still there, and rising sea levels are likely to only affect these areas more. For my project, I sought to see just how badly this less affluent population would be affected, and determine ways to prepare for the inevitability of rising sea levels and the increasing flooding that will accompany it.

II. Datasets

The following datasets were retrieved through various APIs.

1. Boston 1ft. Contours
2. Boston Neighborhood Borders
3. Boston Average Household Income by Neighborhood
4. Boston Fire Hydrants
5. Boston Flood Estimates

The next datasets were created programatically using the datasets above.

1. Hydrants by Neighborhood – This data, given in geojson format, provided the locations of fire hydrants in Boston by Neighborhood.
2. Estimated Neighborhood Coverage – This data, given in json format, provided an estimation for each neighborhood of the percentage of the neighborhood affected by flooding.
3. Average Neighborhood Elevation – This data, given in json format, provided the average elevation for each neighborhood in Boston.

III. Methods

Part 1: Income correlation

For this part of my project, I wanted to see if there was any correlation between Income, and likelihood of being affected by flooding; as well as income and geographical elevation. To begin I retrieved the Contours, Neighborhood borders, and household income datasets. To find the average elevation of each neighborhood, I first iterated over every contour in the dataset. For each contour I used the geospatial library shapely to check and see if there was any overlap between the contour and every neighborhood. If there was, I saved the overlapping section and labeled it with the name of the neighborhood that it overlapped. After this was complete, I iterated over the now sorted contour segments. For each neighborhood, I kept a running total of the total length of the contour segments assigned to the neighborhood. I also kept a running total of the weighted elevations. These values were calculated by multiplying the length of the given contour segment with elevation of the segment. After this was complete, I divided the weighted elevations by the total length of the contours for each neighborhood, giving me an average elevation for the neighborhood. These averages were then normalized using z-scores. Next I normalized the income dataset using z-scores. I then combined the two datasets, and calculated the p-value and correlation coefficient of the data.

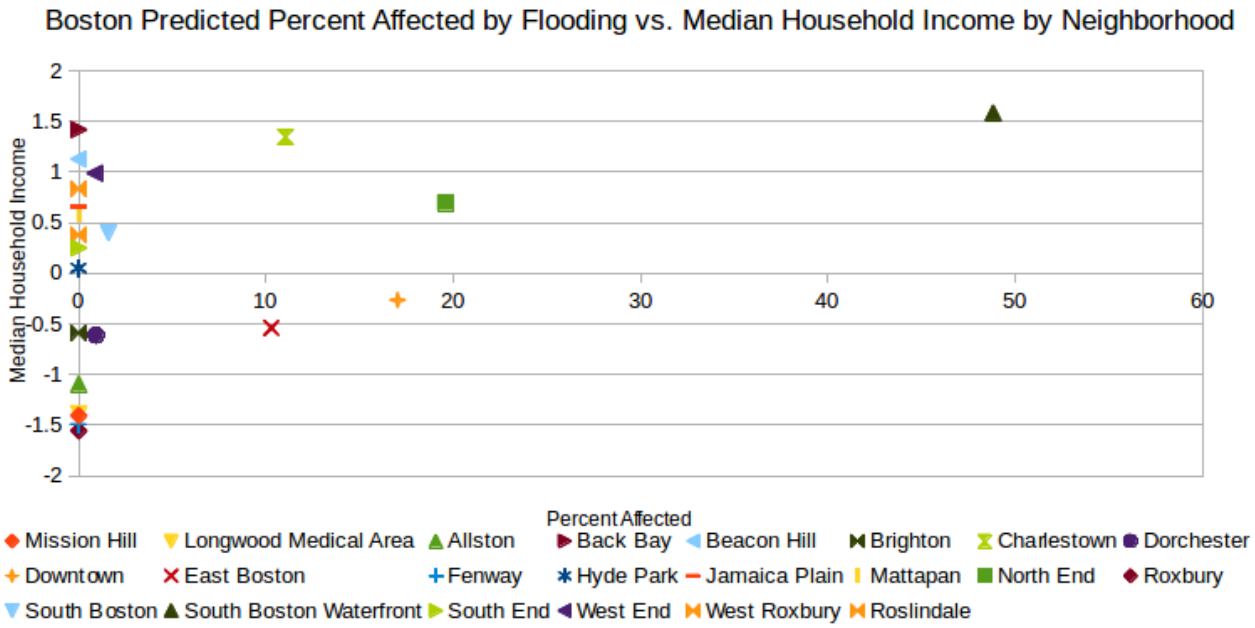
To approximate the percentage to each neighborhood affected by flooding, I used the Boston Fire Hydrants dataset. I first assigned each hydrant to a neighborhood using the geospatial library shapely. Using shapely once again, I determined which hydrants were affected by flooding by checking whether the hydrant overlapped the Boston Flood Estimates dataset. For each neighborhood, I then counted the total number of hydrants in the neighborhood as well as the number of hydrants affected by flooding. I used these numbers to estimate the percentage of the neighborhood affected by flooding.

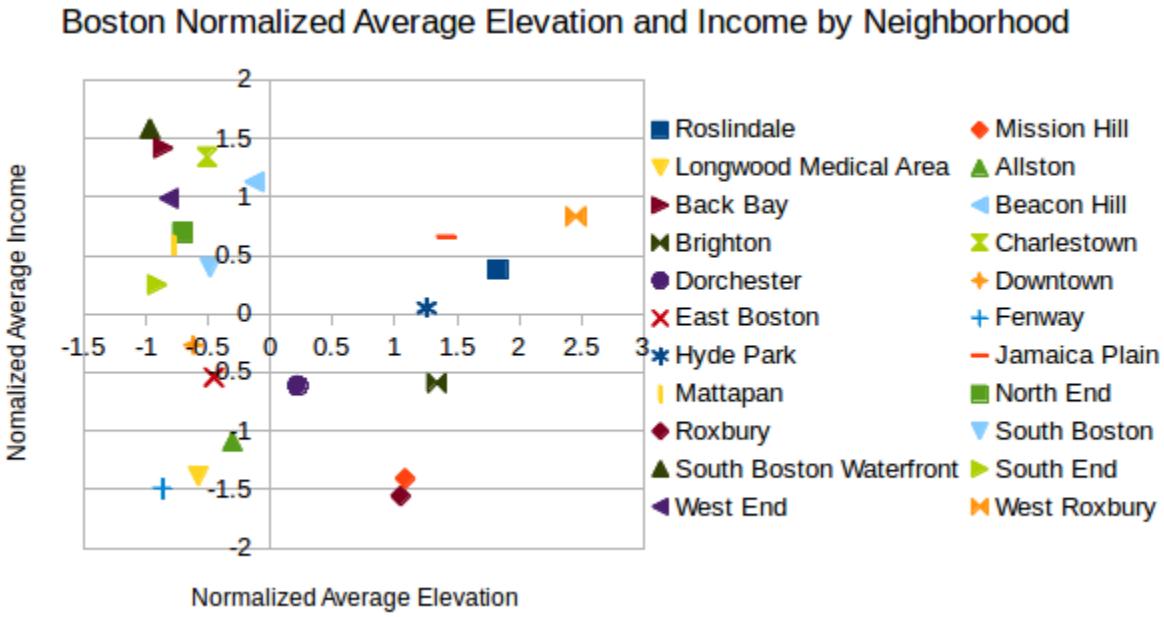
Part 2: Establishing Potential Evacuation Hubs

For this part of my project, I sought to find potential locations for evacuation hubs in the even of extreme flooding. To begin I retrieved the fire hydrants dataset. I then used k-means clustering on the hydrant data to calculate ideal geographic centers for flood relief hubs.

IV. Conclusions and Extensions

There was no clear direct correlation between income and flood risk or elevation. Graphs of both figures can be viewed on page four. The p-values calculated for elevation and flood risk respectively were 0.52 and 0.07. While my analysis didn't find any significant direct correlation, I would be interested to see if there is any indirect risk. I would, for example, be interested in conducting a survey directed towards residents of neighborhoods with high median household income and high flood risk to determine where they would be likely to relocate in the event of extreme flooding. While my k-means algorithm did find geographical centers for evacuation, I would be interested in extending it in the future to calculate optimal routes from these evacuation centers to safety.


Boston Predicted Percent Affected by Flooding vs. Median Household Income by Neighborhood
Median Household Income
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
0
10
20
30
40
50
60
Percent Affected
Mission Hill
Longwood Medical Area
Allston
Back Bay
Beacon Hill
Brighton
Charlestown
Dorchester
Downtown
East Boston
Fenway
Hyde Park
Jamaica Plain
Mattapan
North End
Roxbury
South Boston
South Boston Waterfront
South End
West End
West Roxbury
Roslindale


Boston Normalized Average Elevation and Income by Neighborhood
Nomalized Average Income
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
-1.5
-1
-0.5
0
0.5
1
1.5
2
2.5
3
Normalized Average Elevation
Roslindale
Mission Hill
Longwood Medical Area
Allston
Back Bay
Beacon Hill
Brighton
Charlestown
Dorchester
Downtown
East Boston
Fenway
Hyde Park
Jamaica Plain
Mattapan
North End
Roxbury
South Boston
South Boston Waterfront
South End
West End
West Roxbury